AFRL-IF-RS-TR-2006-253
Final Technical Report
July 2006

NETWORK ANALYSIS AND KNOWLEDGE DISCOVERY THROUGH DNA COMPUTING

JeanSee

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE
ROME RESEARCH SITE
ROME, NEW YORK

STINFO FINAL REPORT

This report has been reviewed by the Air Force Research Laboratory, Information Directorate, Public Affairs Office (IFOIPA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

AFRL-IF-RS-TR-2006-253 has been reviewed and is approved for publication

APPROVED: /s/

THOMAS E. RENZ
Project Engineer

FOR THE DIRECTOR: /s/

JAMES A. COLLINS
Deputy Chief, Advanced Computing Division
Information Directorate

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
JUL 06	Final	Jan 05 – May 06

4. TITLE AND SUBTITLE
NETWORK ANALYSIS AND KNOWLEDGE DISCOVERY THROUGH DNA COMPUTING

5a. CONTRACT NUMBER
FA8750-05-C-0007

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
61101E

6. AUTHOR(S)
Anthony Macula

5d. PROJECT NUMBER
459T

5e. TASK NUMBER
20

5f. WORK UNIT NUMBER
01

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
JeanSee
36 Westview Circle
Geneseo New York 14454

8. PERFORMING ORGANIZATION REPORT NUMBER
N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory/IFTC
525 Brooks Road
Rome New York 13441-4505

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFRL-IF-RS-TR-2006-253

12. DISTRIBUTION AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED. PA#06-548

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The application of computational mathematics and information science to biology has aided in the understanding of biological systems. Today biology can now aid information science. This research activity addresses this new and potentially symbiotic relationship between biology and information. In this report, a biocomputational analysis of a biologically represented network is demonstrated. A report on new DNA aqueous laboratory computing techniques is also given.

15. SUBJECT TERMS
DNA, Molecular Computing, DNA Computer, Knowledge Discovery, Network Analysis

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Thomas E. Renz
U	U	U	UL	24	19b. TELEPONE NUMBER (*Include area code*)

Table of Contents

List of Figures

1. Summary

The application of computational mathematics and information science to biology has aided in the understanding of biological systems. Today biology can now aid information science. This research activity addresses this new and potentially symbiotic relationship between biology and information. In this report a biocomputational analysis of a biologically represented network is demonstrated. A report on new DNA aqueous laboratory computing techniques is also given.

2. Introduction

Classical group testing (CGT) is a widely used biotechnical technique to identify a relatively few number of distinguished objects when the presence of any one of these distinguished objects in a pool produces an observable result. This report describes the algorithmic development of a variant of the classical group testing paradigm called group testing on graphs (GTOG.) The difference between these two abstract models is that GTOG detects edges in a network as opposed to vertices. DNA laboratory techniques and statistical factor analysis are combined to identify covert structures in a network with edges represented by DNA duplexes.

In the laboratory, a variety of methods have been used to detect and modify double-stranded DNA so that it can be used to perform computations and convey information. DNA strands have been successfully tagged with 4 different small molecules, Alexa fluor 488, Bodipy FL, digoxigenin and biotin that enable molecular reading. The first two molecules are fluorescent and reading protocols employing antibodies and binding proteins that recognize these molecules have been tested. These protocols give means of tagging and reading DNA for aqueous computations.

3. Methods

3.1 Algorithmic methods for DNA network

Throughout this paper, all simple lower case Roman variables are non-negative integers. Given set S , $|S|$ denotes its cardinality. Let $[n]=\{1,2,...,n\}$ represent a finite set with n elements which is called the *population*. A k -set in $[n]$ is a k element subset. Let $\Gamma=\{C_1,...,C_d\}$ be an unknown collection of d *disjoint* edges in $[n]$. We refer to Γ as the set of *covert edges*. In GTOG problem, the goal is to identify at least one member of each covert edge in Γ by performing certain 0,1 tests on subsets from $[n]$. A test, $P\subseteq[n]$, is said to be positive if and only it completely contains an edge.

In our application, the graph consists of 33 nodes represented by 33 distinct DNA strands designed (specifically for this proof of principle study) by modifying a SynDCode generated DNA code [1]. A DNA code of size n is a collection of DNA strands with the property that no two strands (including copies of identical strands) have a free energy of duplex formation below a certain threshold. This free energy threshold is selected to ensure that no two strands will hybridize at prescribed experimental conditions (e.g., temperature, ionic concentration.) A complemented DNA code of size 2n consists of n complementary pairs with the property that no two non-complementary pairs (including copies of identical strands) have a free energy of duplex formation below a certain threshold. A complemented DNA code is also known as a DNA tag-antitag system. The graph tested has its vertex set the strands {A,B,...AI,AJ}-{B,H,AE} with its edges depicted in Figure 1. Desired edges are Watson-Crick duplexes and covert edges are cross-hybridized duplexes.

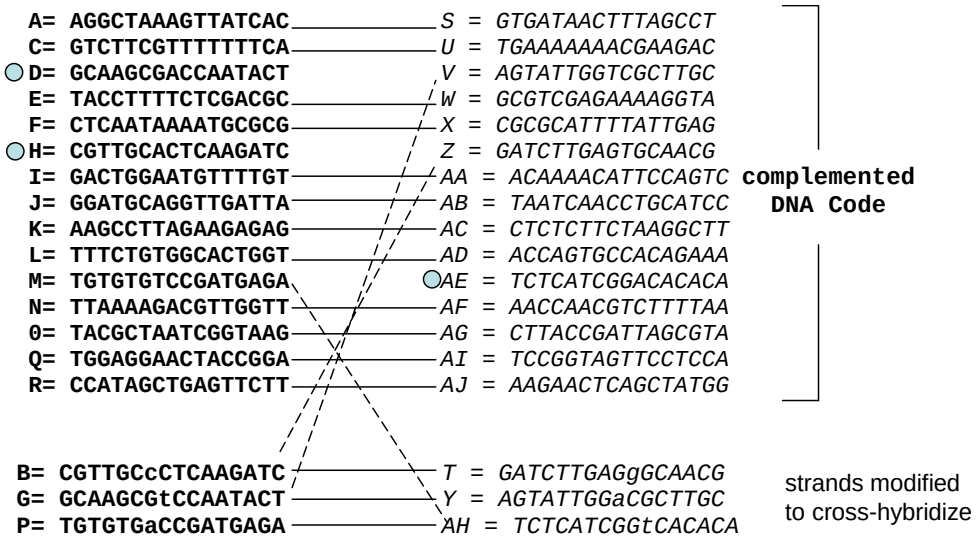


Figure 1: DNA Network

The strands (B,G,P,T,Y and AH) excluded) were generated by SynDCode [1]. In this complemented DNA code, no non-complementary pair has a ΔG of duplex formation less than -8KCal/Mol at 37C. In addition, no non-complementary pair has a secondary structure with more than five 2-stems, or three 3-stems, or two 4-stems. Moreover, since these conditions hold for potential interactions between identical strands, these thresholds also hold for all possible single stranded hairpin structures. Strands D, H, and AE were deleted from the collection. With these stands deleted, exactly three CH duplexes exist at 37C, namely B:Z, G:V, and M:AH. Proper edges are indicated by solid lines and covert edges by dashed lines.

Edge tests for this for this graph consisted of a collection of strands that did not contain a pair of complementary strands (i.e., the desired edges). Thus each edge test corresponds to an independent set in the desired (or assumed) graph. This edge test was then treated with Sybr Green I. This nonsymmetrical cyanine dye is known to produce exponentially greater fluorescence when bound to duplex DNA as opposed to single-stranded DNA. The pools were constructed in a pseudo-random method that is a heuristic modification to a purely random method that can be applied to standard and non-complemented DNA codes.

3.2 Laboratory method for labeling DNA

DNA labeling was accomplished primarily using different types of polymerase chain reaction (PCR) conditions. In general, the multiple cloning region of pBluescript II plasmid with the primers from the -20 and reverse primer from M13 was used in PCR which amplifies the multiple cloning region of this plasmid to form a ~250bp DNA fragment.

Oligonucleotide	Sequence
5'PCRT1T2L	TTGTAAAACGACGGCCAGTGGATCC CCAAACCTCCACTTT CCAAC ACACAACTCC
5'PCRT1F2L	TTGTAAAACGACGGCCAGTGGATCC CCAAACCTCCACTTT CCAACCCTACACCAC
5'PCRF1T2L	TTGTAAAACGACGGCCAGTGGATCC CAACCAACCACTCTA CCAACACACAACTCC
5'PCRF1F2L	TTGTAAAACGACGGCCAGTGGATCC CAACCAACCACTCTA CCAACCCTACACCAC
C(T1)	GTTGGAAAGT GGAGGTTTGG
C(F1)	GTTGGTAGAG TGGTTGGTTG
T2RT3L	TCCACAATCA CCTTTCCTCC
T2RF3L	TCCACAATCA TCACACACAC
F2RT3L	CTACACCTTT CCTTTCCTCC
F2RF3L	CTACACCTTT TCACACACAC
C(T2)	TGATTGTGGA GGAGTTGTGT
C(F2)	AAAGGTGTAG GTGGTGTAGG
T3RT4L	ATCACCTCAT CCTCACTCTC
T3RF4L	ATCACCTCAT CACCTCTCTC
F3RT4L	ACACACAATT CCTCACTCTC
F3RF4L	ACACACAATT CACCTCTCTC
C(F3)	AAT TGT GTG TGT GTG T GTG A
C(T3)	ATGAGGTGAT GGAGGAAAGG
T4RT5L	ACTTCCTTCA TCTCCTCTCC
T4RF4L	ACTTCCTTCA TTTCCACCAC
F4RT5L	ACTTCTTCCA TCTCCTCTCC
F4RF5L	ACTTCTTCCA TTTCCACCAC
C(T4)	TGAAGGAAGT GAGAGTGAGG
C(F4)	TGGAAGAAGT GAGAGAGGTG
T5R3'PCR	ACTCAAAACC A AGC TTC ATG GTC ATA GCT GTT TCC
F5R3'PCR	CTCAACACAT A AGC TTC ATG GTC ATA GCT GTT TCC

C(T5)3'PCR-BIOTIN	/5Bio/GGAAACAGCTATGACCATGAAGCTTGGTTTTGAGT GGAGAGGAGA
C(F5)3'PCR-BIOTIN	/5Bio/GGAAACAGCTATGACCATGAAGCTT ATGTGTTGAG GTGGTGGAAA
5'PCR (M13-UP BamHI)	TTGTAAAACGACGGCCAGTGGATCC
3'PCR (M13-RP HindIII)	GGAAACAGCTATGACCATGAAGCTT
AF5'PCR (M13-UP BamHI)	/5Alex488N/TTGTAAAACGACGGCCAGTGGATCC
Biotin3'PCR(M13-RPHindIII)	/5Bio/GGAAACAGCTATGACCATGAAGCTT

Figure 2: Oligonucleotide sequences for construction of DNA library

3.3 Laboratory methods for DNA network pooling

Pools consisted of 0.5 µL of each DNA strand, 5 µL 25 mM $MgCl_2$. 5 µL reserved for SYBR Green I master mix distilled deionized H_20 to ensure the final volume of pool is 50 µL. The ddH_20 was added to tubule, then MgCl2, and each strand that would be included. The tubule was centrifuged, placed in a PCR machine for the program listed below, SYBR master mix was then added, and tubules were agitated, centrifuged, and then transferred to 96-well plate. The mix was run in a Real Time PCR thermocycler over a temperature range of 30-60 ºC. The program typically took 28 min. The covert edges were then detected by increase in fluorescence.

3.4 Laboratory methods for DNA Library construction and analysis

From DNA sequences designed by SynDCode, several DNA oligonucleotides were used to create a library using parallel overlap assembly [2,3]. The DNA sequences are given in Figure 2.

These sequences were initially mixed as follows. In one tube, 0.5nmoles of 5'PCRT1T2L, 5'PCR F1T2L, 5'PCRT1F2L, 5'PCRF1F2L, T2RF3L , T2RT3L, F2RF3L, F2RT3L with 1nmole of C(F2), C(T2), C(T1), and C(F1) were combined. In another tube, 0.5nmoles of F4RF5L, T4RT5L, F3RF4L, T4RF4L, F3RT4L, F4RT5L, T3RF4L, T3RT4L with 1nmole of C(F3), C(T3), C(F4), and C(T4) were combined. These tubes were heated in a boiling water bath for 10 minutes and then allowed to cool slowly to room temperature. These annealed oligonucleotides were treated with polynucleotide kinase and then ligated for 15 minutes at room temperature. Next, a portion of the second tube was combined with a tube containing 1nmole of T5R3'PCR, F5R3'PCR, C(T5)3'PCR-BIOTIN, C(F5)3'PCR-BIOTIN oligonucleotides that had been annealed and treated with polynucleotide kinase. These combined oligonucleotides were ligated and then combined with the first tube and ligated. Portions of these samples were then amplified with PCR using the 5'PCR and 3'PCR primers and analyzed on polyacrydamide gels. For cloning, we cut the PCR products with BamHI and HindIII and ligated the fragments into a similarly cut pBluescript SKII vector and transformed into E. coli cells. Isolated plasmids were sequenced using automated sequencing reactions and separated on an ABI Model 310 DNA sequencer.

4. Results and Discussion

4.1 Random design model

Let $0 \le p \le 1$. A random pool (or row of a random matrix) in $[n]$ can be described as an n -sequence of i.i.d Bernoulli trials $X = X_1,...,X_n = (X_j)$ where each X_j is 1 with probability p .

Throughout this paper p is reserved and assumed to be $p = \Pr ob(X_{i,j} = 1)$.

Then given n, p, k, d and a fixed $\Gamma = \{C_1,...,C_d\}$ as above, we define the binary random variable Y to be 1 if the pool X is positive. Given N random pools, then for each i in $1 \le i \le N$, then $X = (X_{i,j})$ where $X_{i,j}$ is the ij^{th} entry of the random matrix. With Y_i corresponding to X , we have the sequence $X = Y_1,...,Y_N = (Y_i)$ where the random variable Y_i is 1 if and only if the pool X is positive. In other words,

$$Y_i = 1 \Leftrightarrow \text{there is } C_l = \{j_1,..., j_k\} \in \Gamma \text{ such that } X_{i,j_1} = ... = X_{i,j_k} = 1$$

Clearly, $Y_i^{'}s$ are i.i.d. It is straightforward to verify that for a given $\Gamma = \{C_1,...,C_d\}$ as described above that $\Pr ob(Y_i = 0) = (1 - p^k)^d$. We call Y the *output vector*. See Figure 3.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	Y
pool 1	1	1	1	1	0	0	1	1	0	0	1	0	0	1	0	0	0	0	0	0	0	1	1	1	0	1
pool 2	0	1	1	1	1	1	0	0	1	0	1	0	1	0	1	1	0	0	1	0	0	0	0	0	0	0
pool 3	0	1	1	1	1	0	1	0	0	1	1	0	0	1	1	1	1	1	0	1	1	0	1	1	0	0
pool 4	1	0	1	1	0	0	0	1	0	0	1	1	1	0	0	1	0	1	1	1	1	1	1	1	0	1
pool 5	0	0	0	0	1	1	1	0	1	0	1	0	1	1	0	1	1	1	1	0	1	1	0	0	0	0
pool 6	1	0	1	0	0	1	1	0	0	0	1	1	0	1	0	0	1	1	1	0	1	1	0	1	1	0
pool 7	1	1	0	0	1	1	1	0	1	0	1	0	1	1	0	0	1	1	1	1	0	0	1	0	0	1
pool 8	0	0	0	1	1	0	1	0	0	0	0	0	0	1	1	1	1	0	0	0	0	1	0	0	1	0
pool 9	1	0	1	0	1	1	1	0	0	0	1	1	1	0	0	0	0	1	0	1	1	1	1	0	0	1
pool 10	1	0	1	1	1	1	1	0	0	1	0	0	0	1	0	1	0	1	1	0	1	1	0	1	1	0
pool 11	0	0	0	0	0	0	0	1	1	1	0	1	0	1	0	0	0	0	0	0	0	1	0	0	1	0
pool 12	0	0	0	1	1	1	0	0	0	1	0	1	0	1	0	1	0	0	1	1	1	1	1	1	0	0
pool 13	1	1	0	0	1	0	0	0	0	1	0	1	0	0	0	0	1	1	0	0	1	1	1	0	0	1
pool 14	1	1	1	0	1	1	1	0	0	0	1	0	0	0	0	0	1	0	1	0	1	0	0	1	0	1
pool 15	1	0	0	1	0	0	0	0	0	0	1	1	0	0	0	1	0	0	1	0	0	1	1	0	1	0
pool 16	1	1	0	0	1	1	0	0	1	1	0	0	1	0	1	1	1	0	1	0	1	0	1	1	0	1
pool 17	0	1	1	1	1	1	1	0	1	0	1	0	0	1	1	1	0	0	1	0	1	0	0	0	1	0
pool 18	0	0	1	0	1	0	1	0	0	0	1	0	0	1	1	1	0	1	0	0	1	0	0	0	1	0
pool 19	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0
pool 20	1	1	1	0	1	0	0	1	0	1	0	1	0	0	1	0	0	0	0	0	0	0	1	1	0	1
pool 21	0	0	1	1	1	1	1	0	1	0	0	0	0	1	1	1	0	0	0	0	0	1	1	0	0	0
pool 22	0	0	0	0	0	0	0	1	1	1	0	1	1	0	1	1	0	0	0	0	1	0	0	1	1	1
pool 23	0	0	1	1	0	0	0	0	0	1	0	0	1	0	0	1	1	1	1	1	0	0	0	0	1	0
pool 24	1	0	1	1	1	0	1	1	0	0	0	1	0	0	0	0	0	1	0	1	1	0	0	0	1	0
pool 25	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	1	0
pool 26	1	1	1	0	0	1	1	0	1	0	1	0	0	0	0	0	0	1	0	1	0	1	1	0	1	1
pool 27	0	0	1	1	1	0	1	1	0	0	0	0	1	0	0	0	0	1	1	1	1	0	0	0	0	0
pool 28	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	1	0	0	1	1	0	0	0	1	0
pool 29	0	1	0	0	0	1	0	1	1	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0

Figure 3: A random design

Here k=2, d=3 and Γ={{1,2},{12,13},{24,25}}. Pool i is positive if and only if row i has 1s in columns 1 and 2, columns 12 and 13, or columns 24 and 25.

4.2 Interpreting output

A combination of the random matrix $(X_{i,j})$ and its associated output vector (Y_i) is used to define a new matrix $(Z_{i,j})$. For given n, p, k, d , $\Gamma = \{C_1,...,C_d\}$, and an instance $(X_{i,j})$ as described above, suppose we have the resulting (Y_i). For each $(i, j) \in N \times [n]$, we define the random variable $Z_{i,j} = X_{i,j} + Y_i + 1 \pmod 2$. For each $j \in [n]$, the sequence $Z = Z_{1,j},...,Z_{N,,j} = (Z_{i,j})$ consists of i.i.d Bernoulli variables. However, for $j_1 \neq j_2$ it is not always the case that Z_{i,j_1} and Z_{i',j_2} are independent or identically distributed and it is this difference that we exploit. Clearly the distribution of $Z_{i,j}$ depends on the given j. Consider an instance of a random matrix $(X_{i,j})$ as j column vectors, then the matrix $(Z_{i,j})$ results from adding the complement of output vector $\bar{Y}$ to each column of $(X_{i,j})$. The binary matrix is constructed from $(X_{i,j})$ and Y and shown in Figure 4.

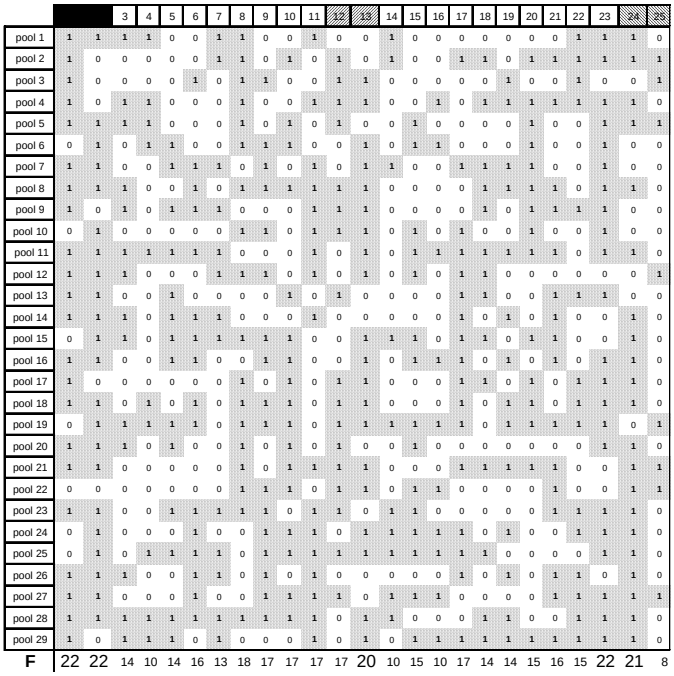


Figure 4: Adjusted random matrix $(Z_{i,j})$

Matrix $(Z_{i,j})$ constructed from $(X_{i,j})$ and $\bar{Y}$. The weights of the columns are given at the bottom. If the threshold is $t = 0.63$, then $\lceil Nt \rceil = 19$. With this value of t, the columns taken are $j = 1, 2, 13, 23, 24$. Notice that all but $j = 23$ is an element of a covert edge and each covert edge has been hit at least once. Also the

values of $p = .467$ and $t = .63$ where chosen to minimize N when 50% of the $\bigcup_{l=1}^{d} C_l$ is expected to be identified and the error rate $(\Pr ob(T \mid G))$ is 0.04.

For each $j \in [n]$, we define $F_{N,j} = \sum_{i=1}^{N} Z_{i,j}$. Then for each $j \in [n]$, $F_{N,j}$ gives the number of $1s$ in, or the *weight* or *frequency* of, the j^{th} column the in matrix $(Z_{i,j})$. Given a $N \times n$ matrix $(Z_{i,j})$ and an a *threshold* t, take, as the objects of the search, all columns of $(Z_{i,j})$ whose weight is at least Nt . See Figure 4. In Section 4.4, the method of choosing the values of p and t so as to minimize the number of pools N is given.

4.3 Distributions, parameters and design functions

For n, k, d and $\Gamma = \{C_1, ..., C_d\}$ as described above, let $Bad = \bigcup_{l=1}^{d} C_l$ and $Good = [n] - Bad$. For $j \in Good$, let p_g^j be the probability that $Z_{i,j} = 1$ and let p_b^j be the probability that $Z_{i,j} = 1$ if $j \in Bad$. It is straightforward to verify that $p_g^{j_1} = p_g^{j_2}$ if $j_1, j_2 \in Good$ and because Γ is a disjoint family we have that $p_b^{j_1} = p_b^{j_2}$ if $j_1, j_2 \in Bad$. Thus we let p_b and p_g respectively denote p_b^j and p_g^j. Below we compute the values of p_b and p_g. These values are different and it is this difference that allows us to distinguish between the $Good$ and the Bad .

For $j \in Bad$, we let $b_{0,0}$ be the probability that $X_{i,j} = 0$ and $Y_i = 0$ and $b_{1,1}$ be the probability that $X_{i,j} = 1$ and $Y_i = 1$. Then $p_b = b_{0,0} + b_{1,1}$. For $j \in Good$, we let $g_{0,0}$ be the probability that $X_{i,j} = 0$ and $Y_i = 0$ and $g_{1,1}$ be the probability that $X_{i,j} = 1$ and $Y_i = 1$. Then $p_g = g_{0,0} + g_{1,1}$. For $j \in Bad$, we let $b_{1,0}$ and $b_{0,1}$ respectively be $\Pr ob(X_{i,j} = 1 \ , \ Y_i = 0)$ and $\Pr ob(X_{i,j} = 0 \ , \ Y_i = 1)$. Similarly, for $j \in Good$, let $g_{1,0}$ and $g_{0,1}$ respectively be $\Pr ob(X_{i,j} = 1 \ , \ Y_i = 0)$ and $\Pr ob(X_{i,j} = 0 \ , \ Y_i = 1)$. As was noted in Section $\Pr ob(Y_i = 0) = (1 - p^k)^d$. It is straightforward to see that since Γ is a disjoint family that $b_{0,0} = (1-p)(1-p^k)^{d-1}$ and $g_{0,0} = (1-p)(1-p^k)^d$. Thus $p_b - p_g = 2(1-p)p^k(1-p^k)^{d-1}$.

For each $j \in Bad, Good,$ we have that $F_{N,j} = b(N, p_b)$, $b(N, p_g)$ respectively where $b(N, p)$ is the binomial distribution. Let $F_{N,j} = F_N^b$, F_N^g for $j \in Bad$, $Good$ respectively. Values for p and a threshold t that discriminate between objects in Bad and $Good$ are desired. Using the normal approximation to the binomial, we can optimize in the following way. Assume F_N^b is normal with $\mu_b = Np_b$ and $\sigma_b = \sqrt{Np_b(1-p_b)}$ and F_N^g is normal with $\mu_g = Np_g$ and $\sigma_g = \sqrt{Np_g(1-p_g)}$. Let t denote a frequency threshold. Let $z_b(N,t)$, $z_g(N,t)$ be the z-score of Nt under F_N^b, F_N^g respectively. Thinking of $z_b(N,t)$ and $z_g(N,t)$ as parameters, we express N and t as a function of p . We have:

$$z_b(N,t) = \frac{Nt - Np_b}{\sqrt{Np_b(1-p_b)}} = \sqrt{N}\,\frac{t - p_b}{\sqrt{p_b(1-p_b)}}$$

$$z_g(N,t) = \frac{Nt - Np_g}{\sqrt{Np_g(1-p_g)}} = \sqrt{N}\,\frac{t - p_g}{\sqrt{p_g(1-p_g)}}$$

Combining these:

$$\frac{z_b(N,t)\sqrt{p_b(1-p_b)}}{t-p_b} = \sqrt{N} = \frac{z_g(N,t)\sqrt{p_g(1-p_g)}}{t-p_g}$$

and

$$t(p) = \frac{p_g z_b(N,t)\sqrt{p_b(1-p_b)} - p_b z_g(N,t)\sqrt{p_g(1-p_g)}}{z_b(N,t)\sqrt{p_b(1-p_b)} - z_g(N,t)\sqrt{p_g(1-p_g)}}.$$

Finally,

$$N(p) = \left(\frac{z_b(N,t)\sqrt{p_b(1-p_b)}}{t(p)-p_b}\right)^2.$$

4.4 Optimization methods

Given n, p, k, d and $\Gamma = \{C_1, ..., C_d\}$ as above and values of z_b ❋z_b ❍❍❑❍❑❑ and $z_g \equiv z_g(N(p), t(p))$, we define the following events for each $j \in [n]$: G is the event that $j \in Good$, B is the event that $j \in Bad$ and T is the event that $F_{N(p),j} \geq \lceil N(p)t(p) \rceil$. We refer to T as the element j attaining the threshold. We have that $\Pr ob(B) = \frac{kd}{n}$, $\Pr ob(G) = 1 - \frac{kd}{n}$, $\Pr ob(T \mid B) = \Pr ob(N(0,1) \geq z_b)$ and $\Pr ob(T \mid G) = \Pr ob(N(0,1) \geq z_g)$. Then the probability that an element j is a member

of some complex given that it has attained the threshold is:

$$\Pr ob(B \mid T) = \frac{\Pr ob(T \mid B)\Pr ob(B)}{\Pr ob(T \mid B)\Pr ob(B) + \Pr ob(T \mid G)\Pr ob(G)}$$

$$= \frac{\Pr ob(N(0,1) \geq z_b)\frac{kd}{n}}{\Pr ob(N(0,1) \geq z_b)\frac{kd}{n} + \Pr ob(N(0,1) \geq z_g)(1-\frac{kd}{n})}$$

$$\sim \frac{\Pr ob(N(0,1) \geq z_b)}{\Pr ob(N(0,1) \geq z_b) + \frac{n}{kd}\Pr ob(N(0,1) \geq z_g)} \text{ when } n \gg k, d.$$

The expected number of objects identified in a given complex is $k \Pr ob(T \mid B) = k \Pr ob(N(0,1) \geq z_b)$. The expected number of objects misidentified is $(n-kd)\Pr ob(T \mid G) = (n-kd)\Pr ob(N(0,1) \geq z_g)$. We consider six special cases:

1. Suppose that, on average, 1 object of each complex is to be identified. In this case, $\Pr ob(N(0,1) \geq z_b) = \frac{1}{k}$ and $\Pr ob(B \mid T) \sim \frac{1}{1+\frac{n}{d}\Pr ob(N(0,1) \geq z_g)}$. Thus to be sure that $0.5, \frac{k}{k+1}$ and $\frac{d}{d+1}$ of the objects attaining the threshold are actually members of a complex, then $\Pr ob(N(0,1) \geq z_g)$ must be $\frac{d}{n}$, $\frac{d}{kn}$ and $\frac{1}{n}$ respectively.

2. Suppose that, on average, half of each complex is to be identified. In this case, $\Pr ob(N(0,1) \geq z_b) = \frac{1}{2}$ and $\Pr ob(B \mid T) \sim \frac{1}{1+\frac{2n}{kd}\Pr ob(N(0,1) \geq z_g)}$. Thus to be sure $0.5, \frac{k}{k+1}$ and $\frac{d}{d+1}$ of the objects that attain the threshold are actually members of a complex, $\Pr ob(N(0,1) \geq z_g)$ must be $\frac{kd}{2n}$, $\frac{d}{2n}$ and $\frac{k}{2n}$ respectively.

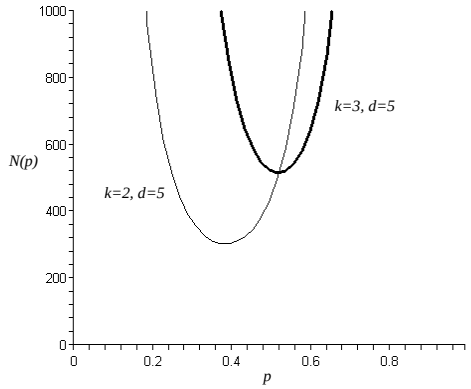


Figure 5: Graph of N(p)

In Figure 5, the graphs of $N(p)$ are displayed. In each case $z_b = -0.253$ and $z_g = 3.090$. Thus the $\Pr ob(T \mid B) = 0.60$ and $\Pr ob(T \mid G) = 0.001$. When $k = 2$ and $d = 5$, the minimum value of $N(p) = 303$ is attained when $p = 0.384$. For this value of p we have that that $0.488 = p_g \le t(p) \le p_b = 0.584$ and $t(p) = 0.577$ Recall that $F^b_{N(.384)} \sim b(303, 0.584)$ and $F^g_{N(.384)} \sim b(303, 0.488)$. Thus for our optimal values, we have:

$$\Pr ob(T \mid B) = \Pr ob(F^b_{N(0.384)} \ge \lceil N(0.384)t(0.384) \rceil = 175) = \Pr ob(N(0,1) \ge -0.253) = 0.60$$

and

$$\Pr ob(T \mid G) = \Pr ob(F^g_{N(.384)} \ge \lceil N(.384)t(.384) \rceil = 175) = \Pr ob(N(0,1) \ge 3.090) = .001.$$

What this indicates is that if $j \in Bad = \bigcup_{l=1}^{d=5} C_l$, then the probability, $\Pr ob(F_{303,j} \ge 175) = 0.6$ and on average, $0.6 \mid C_l \mid = 1.2$. Members of each complex are identified by attaining the threshold of 175. On the other hand, for $j \in Good = [n] - \bigcup_{l=1}^{d=5} C_l$, we have that the $\Pr ob(F_{303,j} \ge 175) = 0.001$. When $k = 3$ and $d = 5$, the minimum value of $N(p) = 515$ is attained when $p = 0.517$. For this value of p we have that that $0.501 = p_g \le t(p) \le p_b = 0.574$ and $t(p) = 0.569$. Since $k = 3$ here, we have on average, $0.6|C_l| = 1.8$ members of each complex identified by attaining the threshold of 296. In both cases $k = 2,3$ and $d = 5$, if $n = 1000$, an average number of ~ 1 member of $Good$ attains the respective threshold and is misidentified as a member of a

complex. Using the equation for Pr*ob*(B | T), we obtain $\Pr ob(B \mid T) \sim \frac{0.6}{0.6+\frac{1}{kd}} = \frac{0.6kd}{0.6kd+1}$.

Thus for $k = 2, \quad d = 5,$ we have that $\Pr ob(B \mid T) = 0.857$ and when $k = 3$ and $d = 5, \quad \Pr ob(B \mid T) = 0.900.$

In general, given n, k, d , z_b and z_g , where z_b is chosen to so that, on average, a certain proportion of each complex identified and z_g is chosen to ensure a certain degree of accuracy given by the conditional probability, we find the value of p that minimizes N . Then we use this optimal value of p to find the value for the threshold *t.*

4.5 Results for the random application to the DNA network

The pools were constructed in a pseudo-random method that is a heuristic modification to a purely random method that can be applied to standard and non-complemented DNA codes. A portion of the abstract random design is depicted in Figure 6. Figure 7 indicates what the theoretical and observed output was from the Real-Time PCR Thermocycler.

	Strands																	Complements																	
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ
Pools																																				
1	1	1	1	*	0	0	1	*	0	0	1	1	0	1	1	1	0	1	0	0	0	0	1	1	0	1	1	1	0	0	*	0	0	0	1	0
2	0	0	0	*	1	1	1	*	1	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	1	0	1	1	1	*	1	1	1	0	0
3	1	1	1	*	1	0	1	*	1	1	0	0	1	1	0	1	0	1	0	0	0	1	0	1	0	1	0	0	1	1	*	0	1	0	1	0
4	1	0	1	*	0	1	1	*	0	0	0	1	0	0	1	0	1	0	0	1	0	1	1	0	0	1	1	1	1	0	*	1	0	1	0	1
5	1	1	1	*	1	0	1	*	0	0	1	0	0	1	1	0	1	0	0	0	0	1	0	1	0	1	1	1	0	1	*	0	0	1	0	1
6	1	1	0	*	0	0	0	*	0	1	1	1	0	1	1	1	0	1	0	0	1	0	1	1	1	0	1	0	0	0	*	0	0	0	1	0
7	0	0	0	*	0	0	1	*	0	1	1	1	0	0	1	0	1	1	1	1	1	1	1	1	0	1	1	0	0	0	*	1	0	1	0	0
8	0	1	0	*	0	0	0	*	1	0	1	0	1	0	0	1	0	1	1	0	1	1	1	1	1	0	0	1	0	1	*	1	1	0	1	0
9	0	1	0	*	1	0	0	*	1	0	1	0	1	0	1	1	0	1	1	0	1	0	0	1	1	0	0	1	0	1	*	1	0	0	1	0
10	0	1	1	*	1	1	0	*	1	0	1	1	1	0	0	1	0	0	1	0	0	0	0	0	1	0	0	1	0	0	*	1	1	0	1	1
11	1	1	1	*	1	0	0	*	1	0	1	1	1	0	0	1	1	1	0	0	0	1	0	1	1	1	0	1	0	0	*	1	1	0	0	0
12	0	0	1	*	0	0	1	*	1	1	1	0	0	0	1	1	1	1	1	1	0	1	1	1	0	1	0	0	0	1	*	1	0	0	0	0
13	1	1	0	*	1	1	1	*	1	0	1	1	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	1	0	0	*	1	1	1	1	1
14	0	1	1	*	0	0	1	*	1	1	1	0	0	1	0	1	0	1	1	0	0	1	1	1	0	1	0	0	0	1	*	0	1	0	1	0
15	0	1	1	*	0	1	1	*	1	0	0	0	0	1	1	1	1	0	1	0	0	1	1	0	0	1	0	1	1	1	*	0	0	0	0	1
16	1	1	1	*	1	0	0	*	1	1	1	1	0	1	1	1	1	1	0	0	0	0	0	1	1	0	0	0	0	0	*	0	0	0	0	0
17	0	1	0	*	1	0	0	*	1	0	0	1	1	1	0	1	1	0	1	0	1	0	0	1	1	1	0	1	1	0	*	0	1	0	0	1
18	1	1	1	*	1	1	1	*	1	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	1	0	1	1	1	*	0	0	0	0	1

Figure 6: DNA Network pools 1-18

Pool i consists of all strands whose column has a 1 in the ith row. Sixty pools in all were used.

	covert edge Contained in Pool?			Observed Fluorescence Pattern	Theoretical Fluorescence Pattern	False Positive	False Negative
	B:Z	G:V	M:AH				
Pools							
1	yes	no	no	positive	positive	no	no
2	no	no	yes	positive	positive	no	no
3	no	yes	yes	positive	positive	no	no
4	no	no	yes	positive	positive	no	no
5	no	yes	yes	positive	positive	no	no
6	no	no	no	negative	negative	no	no
7	no	yes	no	positive	positive	no	no
8	no	no	yes	positive	positive	no	no
9	no	no	no	negative	negative	no	no
10	no	no	yes	positive	positive	no	no
11	yes	no	no	positive	positive	no	no
12	no	no	no	negative	negative	no	no
13	yes	no	no	positive	positive	no	no
14	no	no	no	negative	negative	no	no
15	yes	no	no	positive	positive	no	no
16	no	no	no	negative	negative	no	no
17	yes	no	no	positive	positive	no	no
18	yes	no	yes	positive	positive	no	no
33	no	no	yes	negative	positive	no	yes
58	no	no	yes	negative	positive	no	yes

Figure 7: Sample outcome of DNA pool fluorescence

Both the theoretical and observed outcomes for a sample of pools are given. Note only two pools out of sixty where in error.

A random design of sixty pools was used. In Figure 8, the column frequencies are given. The cutoff was 35. Thus the vertices of the covert edges were identified with 100% accuracy.

29	40	34	29	31	38	28	28	26	29	40	30	34	17	39	29	31	20	26	39	31	29	22	36	32	32	34	31	30	26	43	21	31
A	B	C	E	F	G	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AF	AG	AH	AI	AJ
	*				*					*				*					*				*								*		

Figure 8: DNA network test frequencies

N=60 pools. The threshold is $t = 0.58$, then $\lceil Nt \rceil = 35$. With this value of t, the columns taken are marked with *.

4.6 DNA labeling results

Labeling of the DNA was accomplished by using biotinylated or Bodipy-FL modified reverse primer, Alexa-fluor modified -20 primer, or by using labeled nucleotides modified with either biotin or digoxigenin. These labeled and unlabeled primers were obtained from Integrated DNA Technologies. The amplified DNA was purified using agarose gel electrophoresis and/or the PCR extraction kit (Qiagen). Detection of the fluorescently labeled DNAs was done on a UV transilluminator directly after separation without further staining. Alternatively, the gel was transferred to a nylon membrane and

probed with specific antibodies. We also performed dot blots to confirm the presence (or absence) of tagged DNA using the antibodies. In this case, 0.5-1ul of DNA solutions plus controls were dotted on nylon membrane, treated with UV then probed with the antibodies to one of the specific labels. The antibodies were generally detected with secondary antibodies linked to horse radish peroxidase using a chemiluminescent substrate (Pierce). For the separation of the specifically labeled DNAs in solution, we incubated the antibody with anti-antibody linked agarose or magnetic beads, removed the unbound antibody, then added appropriate amounts of the labeled DNA. These beads contained a complex of tagged DNA and anti-tag antibodies. The anti-antibody beads were then washed and the DNA and anti-tag antibody are removed using boiling water. The presence of the tagged-DNA in each fraction was confirmed using gel separation or dot blots and then anti-tag antibodies followed by anti-antibodies and the chemiluminescent substrate. Light from the substrate was detected by X-ray film and then scanned.

4.7 Results of DNA library construction and analysis

These sequences were initially mixed as follows. In one tube, 0.5nmoles of 5'PCRT1T2L, 5'PCR F1T2L, 5'PCRT1F2L, 5'PCRF1F2L, T2RF3L , T2RT3L, F2RF3L, F2RT3L with 1nmole of C(F2), C(T2), C(T1), and C(F1) were combined. In another tube, 0.5nmoles of F4RF5L, T4RT5L, F3RF4L, T4RF4L, F3RT4L, F4RT5L, T3RF4L, T3RT4L with 1nmole of C(F3), C(T3), C(F4), and C(T4) were combined. These tubes were heated in a boiling water bath for 10 minutes, then allowed to cool slowly to room temperature. These annealed oligonucleotides were treated with polynucleotide kinase and then ligated for 15 minutes at room temperature. Next, a portion of the second tube was combined with a tube containing 1nmole of T5R3'PCR, F5R3'PCR, C(T5)3'PCR-BIOTIN, C(F5)3'PCR-BIOTIN oligonucleotides that had been annealed and treated with polynucleotide kinase. These combined oligonucleotides were ligated and then combined with the first tube and ligated. Portions of these samples were then amplified with PCR using the 5'PCR and 3'PCR primers and analyzed on polyacrydamide gels. For cloning, PCR products were cut with BamHI and HindIII and ligated the resulting fragments were then ligated into a similarly cut pBluescript SKII vector and transformed into E. coli cells. Isolated plasmids were sequenced using automated sequencing reactions and separated on an ABI Model 310 DNA sequencer.

4.8 Results of labeling and detection of tagged DNA

DNA strands were tagged with different modified bases so that they could be read by antibodies. PCR primers that had a reactive amine group incorporated into one of the bases were reacted with succimidyl containing Alex-fluor or Bodipy compounds to form labeled primers. This reaction was successful in producing fluorescently labeled PCR products (Figure 9).

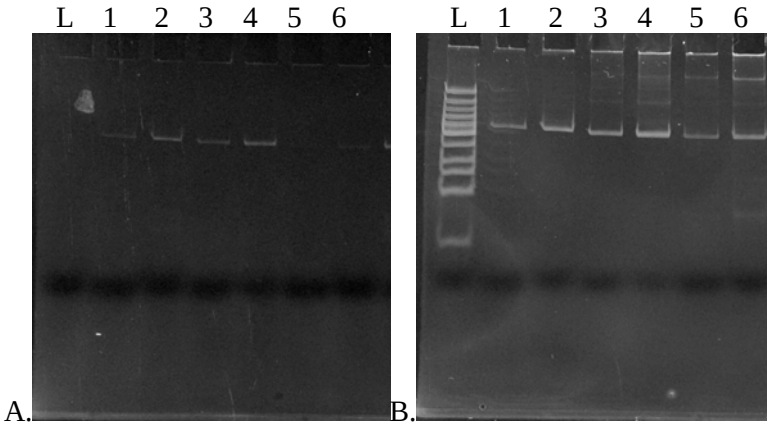


Figure 9: Detection of fluorescently labeled PCR products

PCR made with the Alexa fluor labeled -20 primer (5 and 15ul, lanes 1 and 2, respectively), r the amino-modified with Alexa fluor -20 primer (5 and 15ul, lanes 3 and 4, respectively), and the amino-modified with Bodipy-FL reverse primer (5 and 15ul, lanes 5 and 6, respectively) were separated on a 10% polyacrydamide gel with a low molecular weight DNA marker (L). The DNA was visualized by simply shining UV light (so direct fluorescent bands) (panel A) or through staining of total DNA using ethidium bromide (panel B). Fluorescently labeled DNA with the differently labeled primers and both types of fluorescent molecules were observed.

Antibodies for different tags were tested for their ability to detect the labeled fragments and not any of the other tags. High specificity without cross-reactivity was confirmed by dot blots, loading all of the differently labeled DNA fragments and probing with antibody to Alexa fluor, Bodipy-FL or digoxigenin or with strepavidin that recognizes biotin.

PCR products labeled with Alexa-fluor (AF), Bodipy-FL (BO), digoxigenin (DIG) or biotin were dotted on a nylon membrane and then probed with specific antibodies to Alexa-fluor (far left blot), Bodipy (middle blot) or to digoxigenin (far right blot). Detection of antibody binding was done through anti-rabbit antibodies labeled with alkaline phosphatase and then colorimetric detection of the enzyme. Dark spots indicate enzyme detection of antibody binding. Thus, the antibodies detect only the PCR products labeled with the specific molecule that is their antigen.

4.9 Isolation of tagged DNA using antibody binding

In order to use these tagged DNA fragments in a computation, there is a need to separate DNA that was tagged from unmodified DNA. The general scheme is shown in a diagram in Figure 10. For this, antibodies to those tags bound to beads were used. Protein A agarose beads that bind many types of antibodies including those made against Alexa fluor, Bodipy and digoxigenin were used as were magnetic beads modified with anti-

rabbit antibodies that bind as well to the antibodies to these three tags. For the biotin, strepavidin linked magnetic beads that very effectively bind biotinylated DNA were used.

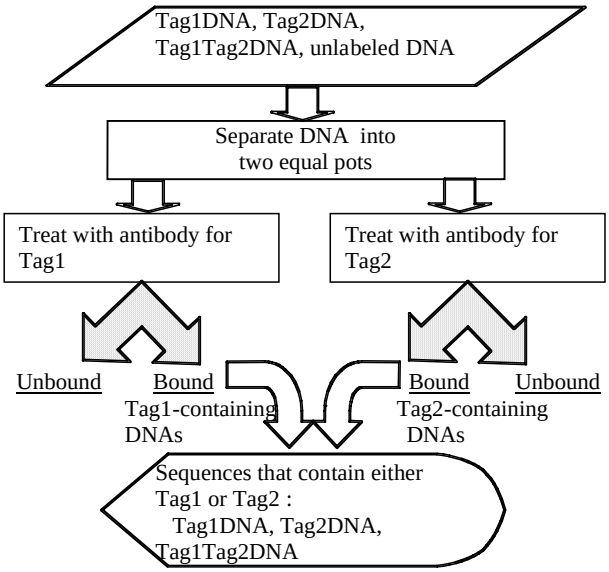


Figure 10: Binding reactions with differently tagged DNA strands

DNA labeled with one or more tags or unlabeled can be separated and put into mixtures in the presence of antibodies for the different tags. Binding of these antibodies and then subsequent separation of the bound material using beads linked with the antibodies will separate out the tagged DNAs. This can be used in a computation to isolate DNA that satisfies the clause Tag1 OR Tag2 (p OR q).

For these reactions, PCR products labeled using one primer with Alexa-fluor and the other primer labeled with biotin were used with one of them bound the DNA with antibodies to the Alexa-fluor linked to magnetic beads with anti-antibodies or with strepavidin linked magnetic beads. The unbound material was then removed after incubation, the beads washed and the bound material isolated by boiling the beads. Detection of the tags in the bound and unbound material was done by either dot blots or by first separating the DNA on a gel, and transfer to a membrane. In both cases, the tags were detected with antibody to Alexa-fluor or with strepavidin that binds to biotin.

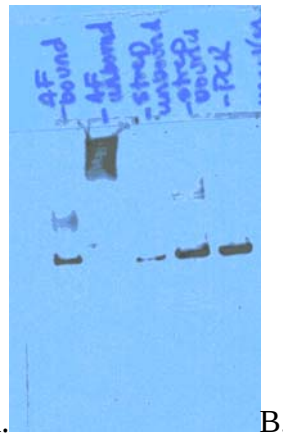
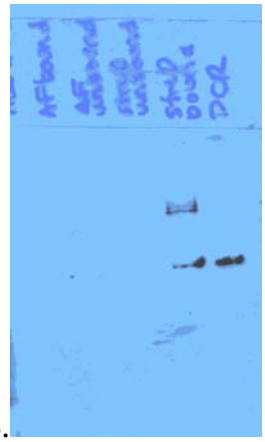
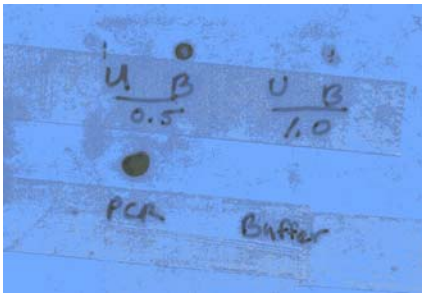
A.  B.  C. 

Figure 11: Separation of labeled DNA using antibodies

DNA labeled with Alexa-fluor and biotin (PCR) was incubated with Alexa-fluor antibodies and the bound (AF bound) and unbound material (AF unbound) was separated on a gel. This DNA was also incubated with strepavidin and the bound (strep bound) and unbound material (strep unbound) was separated on the same gel. See Figure 11, above. Detection using the anti-Alexa fluor antibody, panel A, and streptavidin, panel B, was conducted. Labeled DNA is detected in the bound and some in the unbound fractions. The bands that are higher in the gel than the PCR product is likely to be antibody that may be linked to DNA. In panel C, different amounts of anti-Alex fluor antibody (0.5 or 1.0 ul) was used to see if that changed the binding reaction. Then the bound (B) and unbound (U) material was dotted onto nylon membrane and biotin detected with strepavidin. In both cases, there was no tag detected in the unbound fraction but it appeared that lower amounts of antibody could bind more PCR product to the beads.

4.10 Construction of the DNA library

A 5-site, 2-variable library of sequences using codes designed by SynDCode was constructed by combining the oligonucleotides using a parallel overlap assembly approach. We combined the oligonucleotides in 3 stages and monitored the appearance of appropriate bands during the assembly using polyacrydamide gel electrophoresis (Figure 12).

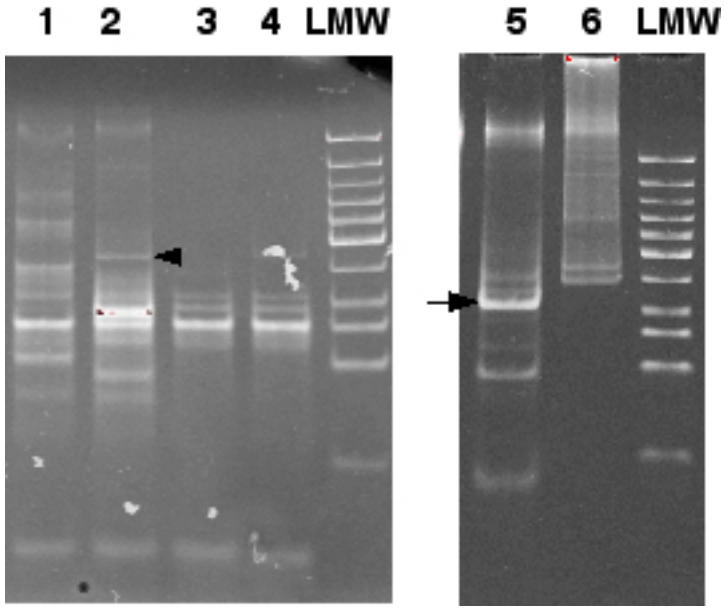


Figure 12: Library preparation, amplification and digestion

A ~175bp band appeared in the final ligation step as expected (Figure11, lanes 2 and 4). PCR amplification of these mixtures using primers at the ends of the library sequences, produced the appropriate sized products (Figure 11, lane 6).

Fragments representing sites 1-2 were combined and ligated and produced primarily fragments at 75bp as expected (lanes 3 and 4) while fragments representing sites 3-4 (lane 1) and sites 3-5 (lane 2) showed major fragments within the mixture at 75bp and ~85bp, respectively. Ligation of fragments representing sites 1-5 produces a ~175bp fragment (marked with an arrowhead in lane 2; also present in lane 4) while ligation of fragments representing parts 1-4 produces a faint slightly smaller fragment (lane 1). LMW represents molecular weight marker bands of size (starting from bottom) 25bp, 50, 75, 100, 150, 200 (slightly brighter), 250, 300, 400, 500, 766bp. PCR of the ligation mixture from lane 2 with primers representing sequences outside the site 1 and site 5 (3'PCR and 5'PCR) produces two bands at 150 and ~175bp (lane 6). When this DNA is cut with BamHI and HindIII (situated just inside the primer sites), a strong ~110bp band is produced (marked with arrow in lane 5). This is the expected size for a band to

represent 5 sites of 20bp each having extra 12 bp to represent the two restriction sites. This digested band was cloned into the vector pBluescript for sequencing.

4.11 Analysis of DNA library

Library strands were cloned to check whether representative sequences were present. For cloning and sequencing of these library DNAs, amplified product was cut with BamHI and HindIII and purified the fragment on a gel (Figure 12, lane 5) where the expected ~110bp product was formed. A number of colonies were obtained and 6 were sequenced. Table 2 contains the sequences of the clones and the site identifications for each. The 6 sequences obtained represented unique different fragments of the predicted library with one clone have a mutation in the T3 site (clone A) and one clone have a single extra base at one end (clone 5). These changes could have been present in the oligonucleotides used to produce the library or could have been introduced by the Taq polymerase that was used to amplify the library before cloning. Finding only these unique fragments representing individually all the possible sites (F or T at each of the 5 sites), is consistent with the hypothesis that we have obtained all 32 individual members of the library during this construction process

Clone name	Clone sequence (site 5->site 1 direction)	Site designations
E	ATGTGTTGAGGTGGTGGAAA- TGAAGGAAGTGAGAGTGAGG- ATGAGGTGATGGAGGAAAGG- AAAGGTGTAGGTGGTGTAGG- GTTGGTAGAGTGGTTGGTTG	F1 F2 T3 T4 F5
G	GGTTTTGAGTGGAGAGGAGA- TGGAAGAAGTGAGAGAGGTG- ATGAGGTGATGGAGGAAAGG- TGATTGTGGAGGAGTTGTGT- GTTGGAAAGTGGAGGTTTGG	T1 T2 T3 F4 T5
A	ATGTGTTGAGGTGGTGGAAA- TGGAAGAAGTGAGAGAGGTG- AATTGTGTGTGTGCGTGTGA- AAAGGTGTAGGTGGTGTAGG- GTTGGAAAGTGGAGGTTTGG	T1 F2 F3* F4 F5 (has a T→C mutation in F3 sequence)
5	ATGTGTTGAGGTGGTGGAAA- TGAAGGAAGTGAGAGTGAGG- ATGAGGTGATGGAGGAAAGG- TGATTGTGGAGGAGTTGTGT- GTTGGAAAGTGGAGGNTTGA	T1 T2 T3 T4 F5 (has an extra A at the end of T1)
4	GGTTTTGAGTGGAGAGGAGA- TGGAAGAAGTGAGAGAGGTG- AATTGTGTGTGTGTGTGTGA- TGATTGTGGAGGAGTTGTGT- GTTGGTAGAGTGGTTGGTTG	F1 T2 F3 F4 T5
F	ATGTGTTGAGGTGGTGGAAA- TGGAAGAAGTGAGAGAGGTG- ATGAGGTGATGGAGGAAAGG- TGATTGTGGAGGAGTTGTGT- GTTGGAAAGTGGAGGTTTGG	T1 T2 T3 F4 F5

Figure 13: Sequences of individual clones from DNA library

5. Conclusions and future work

5.1 DNA networks

The fact that simple network motifs can be discovered using DNA has been demonstrated. Future work entails algorithm development to find more complex motifs.

5.2 Tagged DNA

DNA fragment strands were successfully labeled with 4 different molecules, and it was shown that the antibodies to the molecular tags were indeed specific to those tags. The process of perfecting protocols to select tagged DNA using tag antibodies continues. The highest yielding method is expected to be the permanent linkage of the antibody to the magnetic beads. This method should reduce the noise in the detection procedure when the antibody is still present.

5.3 DNA library construction, analysis and selection

A 5-site 2-variable DNA library using sequences generated by SynDCode was created. Statistical sampling from this collection indicates that the library is likely complete as all fragments were unique members of this 32-member library. Work on protocols for detection of specific probe members using complements immobilized on membranes continues.

References

1.SynDCode website: http://s53n101.academic.geneseo.edu/ , M. Bishop, A. Macula, T. Renz

2. Kaplan PD, Ouyang Q, Thaler DS, Libchaber A.(1997) Parallel overlap assembly for the construction of computational DNA libraries. J Theor Biol. 188:333-41.

3. Ouyang Q, Kaplan PD, Liu S, Libchaber A.(1997) DNA solution of the maximal clique problem. Science. 278:446-9.

4. Adleman LM. (1994) Molecular computation of solutions to combinatorial problems. Science. 266:1021-4.

5. Gal, S., Monteith, N., Shkalim, S., Huang, H. and Head, T.: Methylation of DNA may be useful as a computational tool: Experimental evidence, In Proceedings of the Conference on Mathematical Biology and Dynamical Systems**,** in press.

6. Head, T., and Gal, S.: Aqueous computing: Writing on molecules dissolved in water (2006) In Nanotechnology: Science and Computation, J. Chen and N. Jonoska, Eds., Springer-Verlag publishers, Berlin, pp. 321-334..

7. Head, T., Chen, X., Nichols, M.J., Yamamura, M., & Gal, S.: Aqueous solutions of algorithmic problems: emphasizing knights on a 3X3. (2002) In DNA computing. Lecture Notes in Computer Science, N. Jonoska and N.C. Seeman, Eds, Springer Verlag publishers, Berlin, pp. 191-202.

8. Head, T., Chen, X., Yamamura, M. & Gal, S.: Aqueous Computing: a Survey with an Invitation to Participate (2002) Journal of Computer Science and Technology, 17: 672-681.

9. Head, T., Yamamura, M. & Gal, S.: Aqueous computation: Writing on molecules. (1999) Proceedings of the 1999 Congress on Evolutionary Computation, 2: 1006-1010.

10. Pogozelski, W., M. Bernard, S. Priore, A. Macula, Experimental Validation of DNA Sequences for DNA Computing: Use of a SYBR Green I Assay, DNA 11, Lecture Notes in Computer Science, A. Carbone and N. Pierce, Eds., 3892, 248-256. (2006).